THE CHEMICAL BASIS OF HEREDITY

by

Arthur Kornberg, M.D.

Transcription of a Tuesday Evening Series Talk
October 17, 1961

THE CHEMICAL BASIS OF HEREDITY

By Arthur Kornberg, M.D.

Transcription of a Tuesday Evening Series Talk, Oct. 17, 1961

I think most of you in this audience have through experience or intuition, or both, some fundamental grasp of the basic meaning of heredity. I think you know that starting with a sperm and egg, whether the species be a mouse, or a man, or an elephant, that these microscopic cells, quite similar, often indistinguishable in appearance, unite and from this union two cells appear, and from two, four; and from 4, 8; 8, 16; etc., and in some 40 divisions of this kind, the enormously intricate organism containing many trillions of cells appears and we recognize it as a distinct species--whether it be a frog, or a man or a mouse, despite the fact that the microscopic cells from which this organism originated were almost indistinguishable. I think you also know that this adult individual has reproductive facilities that in turn lead to the production of sperm and egg which perpetuate the species. I think you probably, with some thought, realize that within the sperm or egg is some chemical information that dictates--predestines the most subtle detail--of the individual that is formed: the color of the hair, the color of the eyes, the shape of the nose; and you know from elementary biology that this information is contained in the nucleus of the cell. And you probably recall that the structures designated as chromosomes within the nucleus are further the responsible areas for this chemical information.

Now what we can say further is that something you cannot see under the light microscope is contained within these chromosomes, something that we define as "genes." These are best defined as those sub-units, those invisible sub-units within the chromosomes that are responsible for each simple chemical fact that constitutes the adult individual. Now what I should like to talk about is our approach to understanding these phenomena in simple chemical terms and this evening I should like

to tell you what we know, what we'd like to know about the chemical basis for these gross hereditary phenomena. We would like to know how these genes serve not only to give information for the development of the cell, but are self-reproduced so that progeny in subsequent generations will get the same information and perpetuate the species. These genes, we recognise, can be as few in number as the chemical facts that they are going to dictate; as few as 10 in the simplest virus and as numerous perhaps as 10 million in organisms as complicated as ourselves. Now I think these things we know, or have come to appreciate, thinking from biologic or chemical directions.

We can also approach this as some people have recently, from areas as diverse as electronics or physics, and here in Palo Alto and Stanford, which is a young capital of the electronics industry, we are familiar with enormous advances in miniaturising electrical circuits and you know of tiny transistorised radios and motors that are even smaller than your fingernail. A very interesting essay on this subject was written recently by Professor Richard Feynman at Cal Tech, with the challenging title, "There is Plenty of Room at the Bottom," pointing out the vast area still possible in miniaturisation. He offered in this essay or speech that preceded it, perhaps foolishly, to pay $1,000 to someone who would make a motor that was one-sixty-fourth inch in diameter; and I am told that very soon after someone made it and he had to pay the money. He also pointed out that miniaturization in coding information had hardly progressed at all. It is true there are devices which, he stated, permitted the reproduction of the Pledge of Allegiance on a pinhead and by now I imagine that there is room on such a pinhead for other oaths. But what he went on to say is that this is terribly primitive, that it should be possible with known physical principles and certain suggested techniques, to reduce the Encyclopedia Britannica, all twenty-four volumes of it, to the size of a pinhead. He pointed out, and I will point out to you, that the techniques for doing this are still decades away. But even if this were accomplished--and this represents a demagnification of some 25,000 diameters--

even if this were accomplished, the feat would still pale, by comparison, with what nature has accomplished in reducing genes into chemical language. A dot at the end of this pinhead-size Encyclopedia Britannica would cover an area that can encompass a thousand different atoms, each of which carries information much as our alphabet does. Now this calculation can be made in a variety of ways. You can proceed from this and point out that knowing the mass of genetic material in chromosomes of all the sperm and egg that have gone to make the three billion people on the face of the earth, you would still have a volume not much bigger than this pinhead. This is the feat of chemical language as it has developed in nature. And to be sure that I have made this analogy clear, I want to point out that when I gave this kind of comparison in a talk in Washington this spring, the talk was reported under this headline, " 'Miniaturization May Make Possible Library in Thimble,' says scientist." Now if there are any librarians in the audience, please don't take this seriously.

We see then that the gene can best be expressed as an encyclopedia, and I hope I am not boring you to say that billions of volumes of information can be compressed into these pinhead-size capsules if need be, This encyclopedic gene has information in it which dictates all the development of the cell, in terms of the proteins, the enzymes that are going to be formed, and it must also have the capacity to be copied faithfully so that the progeny of this individual will have the same encyclopedic detail for all the elaborate processes that characterise cellular chemistry.

In pursuing this subject from now on, I would like to focus on three questions. The first is: How have we come to recognise that the genetic substance is DNA, deoxyribonucleic acid, and what is DNA? Secondly, we would like to ask: How is DNA translated into proteins, proteins and enzymes (which are proteins) being the elements in the cell that govern all the chemical machinery and constitute the structure of the cell? And finally: How is DNA copied so faithfully that the race is preserved and the progeny that issue from this cell or individual are virtually identical?

Now for the first question: How have we come to recognise that DNA is the

genetic substance? In this, in effect, encompasses much of the progress in biology and biological chemistry in the last century. One direction we can take is from the work of a young Swiss chemist, Miescher, about one hundred years ago. He wanted to know what the nucleus of a cell consisted of and so he went to the sperm of herring or of other fish at spawning time, where this material was abundant, extracted it and found that the substance which made up much of the nucleus contained the usual elements that are found in the cell--carbon, nitrogen, hydrogen, oxygen and an element that previously had not been found in organic linkage in nature, phosphorus. This was a phosphorus-rich material. His researches, and subsequent studies, showed that this material could be separated from proteins and other elements in the cell. It was called, after its origin, nucleic acid. It has come to be recognised that there are two kinds of nucleic acid: deoxyribonucleic acid (DNA), which is found only in the nucleus, and another kind, RNA, ribonucleic acid, which is found both in the nucleus and the surrounding cytoplasm.

Now what has occurred over the years is the development of stains that are specific for DNA (first slide). We have in this picture a section across the tip of the root of a growing onion plant. Some of you may recognise that I have borrowed this from a recent Scientific American. What you see in these cells and in the nuclei of the cells are these purple bodies, and these, as you watch them, will divide so that an equal amount is contained in each cell. Here are some of the chromosomal elements, and what has been learned over the years is that stains which are specific for DNA actually characterise these chromosomal elements that are so prominent in cellular reproduction. This is one line of investigation and the evidence, while apparently convincing, is still only circumstantial.

Another line of evidence that is much more compelling developed some twenty years ago from the work of Avery and his associates at the Rockefeller Institute, who exploited earlier observations by some English workers. This has to do with the transformation of bacteria. They observed that if you take a bacterium, the pneumococcus

in this case, which has a characteristic capsular coat, that it is possible to extract a substance from one bacterium, get it, insert it in another and change the coat from the type of cell that you extracted it from and impose that on the recipient. In other words, you can in effect extract a character and insert it into another organism that doesn't have it, and from that point on, this is a permanent genetic character. This has now been done in a great variety of ways. Streptomycin resistance, for example, can be transferred from one organism to a closely related one (in effect a different species) and transform that organism into this new species. Avery and his associates showed that the transforming substance is deoxyribonucleic acid. Dr. Lederberg in the Department of Genetics and his associates are doing some of the most advanced studies with other kinds of bacteria, using these naked genes, in effect, that are isolated from cells inserted into others and convert these recipients into the kind of species from which the DNA was obtained.

Another line of evidence that I'd like to cite (and there are numerous ones, but I shall restrict myself to these few) comes from a study of viruses. Viruses, as you know, are very tiny organisms that can survive in nature by themselves, but must infect a larger cell in order to multiply. Such viruses can infect animal cells, but perhaps the most revealing information about the nature of these viruses comes from studies of infections of bacteria with viruses. This next slide shows a rather artistic conception of the life cycle of one of these bacterial viruses, called a bacteriophage. We have here a picture of this virus attached by its tail to a bacterium. In this case the bacterium is escherichia coli, the common inhabitant of our intestinal tract. The work that Hershey and his associates did about ten years ago shows that the essence of this infection by this bacteriophage was an injection of the DNA from the head of the virus. The outer protein coat--a little protein overcoat that it carries--was left behind in the medium. Simply as a result of this injection of the DNA, the chromosome of the host bacterium sort of dissolves and the cell becomes a factory for the reproduction of virus particles. This happens within

seconds, and within ten minutes these numerous packets of DNA are reproduced. They are then ensheathed in protein overcoats and at twenty-one minutes, 100 or 200 particles emerge. The evidence is most convincing that the protein that surrounds the virus has nothing to do with the infection, that the character of the virus is determined entirely by the DNA which is injected.

There is a variant on this kind of infection in which the bacterial virus doesn't quite take over with as much force and interpolates itself in a sort of saprophytic existence with the genetic material of the bacterial host. This is called lysogenic infection and they live happily together until some moment when a rare event occurs or ultraviolet light is shined on this combination of virus and bacterium and then there is this explosive infection. What Dr. Lederberg showed some years ago is that because of this intimate relation between the virus and the genetic apparatus of the host, these viruses can actually carry with them some of the genes of the bacterium. It can be demonstrated that some of these genetic properties they carry can be transferred to other cells that they infect. Dr Dale Kaiser and Dr. David Hogness in the Biochemistry Department have shown you can do this entire thing with just the DNA itself. You can extract the DNA from the virus, infect the cells with the DNA and then develop the new bacterial particles and the DNA will carry some of these genes from the host bacterium to the cell that it infects.

In short, all of the detailed genetics on bacteria and viruses which have been described over the years, can now be studied with DNA molecules in the test tube. Kaiser and Hogness are actually shearing these molecules into sub-units and determining which parts of the molecule carry which information as they infect the cells. So we have come to the point now of saying, yes, DNA is the genetic substance.

What <u>is</u> DNA? I think we can say without elaborate formulae that DNA consists of four letters, assuming that DNA is this encyclopedia we've described. The DNA language then is a four-letter alphabet and these four letters will be referred to from now on as A, T, G and C. What does a letter stand for? The letter A (next slide)

stands for a molecule which is represented here as accurately as chemists know how to do it today. These wooden balls designate atoms. A hydrogen atom is white, oxygen is red, blue is nitrogen, black is carbon and the salmon color is phosphorus. There are some thirty-five atoms here. We have a phosphate, actually linked to a sugar group. This blue and black structure is adenine, and adenine is "A" in the DNA alphabet. I should say that this model is about 200 million times expanded over its correct size and probably blown up another couple of times here. In addition to A in the DNA alphabet, there are very similar units which are called guanine, thymine and cytosine. (Next slide)

It is a remarkable fact that in nature, no matter where you find DNA, whether it is in soil, bacteria, in animals or in fish, the amount of A is always equal to the amount of T. The letters in this alphabet show an invariable equivalence of A to T and of G to C. But the ratio of AT pairs to GC pairs does vary. The ratio is constant and characteristic of a given species. This fact was recognized about 1950, but was not readily explained until about three years later, when Watson and Crick came out with their now very famous model for DNA which reconciles this relationship and many other chemical and physical facts about DNA.

They picture (next slide) that DNA is a collection of these letters, running thousands of units long, and is formed as a helical structure. It winds around like a spiral staircase and jutting out at each point are these letters, A, T, G or C. Now if this were the only fact known about it, we might expect that such a spiral would collapse into a ball like twine rumpled up; but what they showed, and the most significant finding, is that there are actually two such helices that wind about each other as shown here and they are held together by the fit of G to C, C to G, A to T--there are chemical entities in each of these letters that make the fit of C to G perfect, but C will not fit with an A nor G with a T.

We have a picture of a DNA molecule as an enormous fiber. This pointer might represent it to give you some feeling for its rigidity. Within it, as this atom model

may suggest to you, we have this winding staircase with steps uniting these two strands, created by A to T, G to C, T, A, C, G, etc. I would remind you that because of expense, we have not made these models many thousands of units long. This is only ten units long, and it's our conception that the simplest virus known to us which may have just a few genes is at least several thousand units long. A human chromosome or the chromosomal material within a cell is at least several billion units long. What does this structure mean? It means that this alphabet spelling TC, CG, AC, etc. in sufficient length will spell a message and that message will lead to a protein or an ensyme.

Now we come to the second question: How is DNA expressed in protein synthesis? How is this structure, spelled in these four letters of A, T, G and C, translated into proteins which govern all the chemical processes in the cell? The difficult thing here--this problem has not been solved--is the translation of this four-letter language into proteins which are made up of twenty letters. Proteins, as you know, consist of amino acids. All proteins in nature virtually are made up of twenty different amino acids. How do you translate a four-letter language into a twenty-letter language? As I have said, this code has not as yet been deciphered and represents one of the major problems in biochemistry, or molecular biology as it is referred to today. Progress of some sort is being made in trying to figure out how within the cell the DNA leads to the synthesis of proteins. And in the forefront of this field are Dr. Charles Yanofsky in Biology and Dr. Paul Berg in Biochemistry.

Without going into detail, and I am afraid even if there were time, I wouldn't have the details for you, we speculate--this is the best word to use--that the path of translation of the DNA information into protein synthesis may take this turn. (Next slide) We might imagine that this saw-toothed little structure in the nucleus of the cell (designated as DNA) is translated by the use of RNA, and I shan't go into the reasons for this, but they are becoming more and more convincing. Through chemical matching the RNA copies the sequence in DNA and then proceeds as a messenger,

going out of the nucleus somewhere into the cytoplasm and seats itself at this locus, which we shall represent as a ribosome, the place in the cell where the proteins are assembled. Attracted to this messenger RNA are now amino acid-bearing units that fit by a decodification, that I have warned you we don't yet understand. But somehow the structure of this RNA attracts specific amino acids at each point and now in this section of the diagram, we see them assembling and getting close. And when they finally unite as a string of amino acids, we then have a protein. This is an artist's conception of a long string of amino acids, which in a three-dimensional picture give you this kind of protein structure. Actually this is a model of myoglobin, the hemoglobin that is found in muscle, as deduced by Kendrew and Perutz in England. It is the peculiar arrangement of the amino acids and then the three-dimensional structure that they result in that is responsible for the property of this molecule in carrying oxygen in the cell.

You might wonder what would happen if there were a mistake in the DNA message. What would happen if you substituted a T for a G or an A for a C? We think we know the result of this kind of error or alteration and we recognize it as disease. Dr. Pauling did the most original and informative early work in this area of so-called molecular diseases. One condition, for instance, is known as sickle cell anemia, which is a very severe and often fatal disease, especially common in colored people. It has been shown that the hemoglobin of these people with sickle cell anemia is defective in only one regard in the chain of three hundred amino acids that constitutes hemoglobin. There is just one mistake, and that single amino acid alteration produces a molecule that cannot carry oxygen so well, and as a consequence there is this fatal disease. To describe this diagramatically (next slide), we imagine then, and I emphasize that we do not know the sequence in the DNA, let alone the particular segment that is responsible, but we can imagine that the DNA has a code of TCT and these three letters are then translated as the amino acid proline, or GAC is translated as the amino acid glutamic acid, If this is true and the three hundred amino acids are

correct, we have hemoglobin A, which is normal. But assume that instead of an A there is a C. Then the sequence GCC spells valine, another amino acid when it replaces glutamic acid we have the sickle cell hemoglobin, hemoglobin S, and a fatal disease.

During the last few years it has become increasingly apparent that extremely complicated diseases can be traced to a lesion so simple as the alteration of a single amino acid in a single protein or a single enzyme. For example, the disease phenylketonuria, which results in a severe mental retardation, has a variety of symptoms; muscular abnormalities, loss of pigment usually seen in the predominance of blondness and blue eyes, inability to walk, tremors, etc.---all due to the lack of an enzyme or functioning enzyme, that will put an oxygen on one place on phenylalenine to convert to tyrosine; and for lack of that, this whole syndrome of symptoms. This book I am holding, which one can barely lift, has 1500 pages, costs $30.00--was published last year and is already out-of-date, describes a great variety of diseases which can now be traced more or less to simple defects of this nature.

Now we come to the last question. Since you all understand how DNA dictates protein synthesis, we appreciate that this is one of the two functions of DNA. (Next slide). DNA, as I have been emphasizing right along, has a dual role. We have talked about the intervention of RNA, which presumably leads to enzymes or proteins which are responsible for cell development and cell function. And this would suffice for the life of this cell, but it would not permit reproduction. For the cell to reproduce we must copy this DNA, we must copy it exactly so that the progeny of this cell will have exactly the same encyclopedic information to result eventually in protein synthesis.

How do we go about figuring out how DNA reproduces or replicates? I would like to describe this in some detail. It concerns laboratory investigations that Dr. Robert Lehman and I initiated in St. Louis and are still carrying on here at Stanford. The first thing one must do is simplify the cell, so to speak. The cell contains thousands

upon thousands of enzymes all doing their highly integrated jobs and standing outside you are aware that something is going on. Things are going in and out, but what takes place in intimate detail is difficult to fathom. It is essential to understand the individual chemical reactions. So you open the cell and separate the reactions that you're interested in, away from all the other things going on within the cell. This is a rather simple way of saying that you must purify out the particular catalyst, the particular enzyme that is responsible for the phenomenon that you wish to understand; in this case, the synthesis of DNA.

Before tackling this problem, in several laboratories including our own, we had first to understand how the building block (an A, G, T or C, for example) is produced, how this collection of some thirty odd atoms comes together. The A is assembled from more primitive units, like carbon dioxide and ammonia and inorganic phosphate. Some years ago we already understood how each of these building blocks is assembled within the cell from studies using enzymes extracted from the cells. Now we were able to tackle the question of how these units are assembled into this enormous DNA molecule. In carrying out these kinds of studies, I should tell you that we look where the light is brightest. We will look in pigeon liver or in extracts of bacteria or the bone marrow of a rabbit. I am enumerating the various sources from which the knowledge concerning the synthesis of these building blocks was derived and now we learn from studies of this kind the most impressive fact--that no matter where you look in nature, the fundamental processes in synthesizing nucleotides or nucleic acids or proteins or carbohydrates are essentially the same. There is a unity of biochemistry that pervades all of nature. And we can look with confidence at a particular process in escherichia coli, this little microbe in the intestinal tract, for the way in which an animal assembles its own nucleic acid and I can tell you that this confidence has almost always been well placed. So in fact we did look into escherichia coli. It is an organism that divides almost every twenty minutes and is a far better place for studying DNA synthesis than cells in

animals or plants, which divide only once every day or so. The mechanism for proceeding with such studies is to grow some bacteria, grind them up, extract their juices and then by one device or another see whether you can measure the conversion of single units into this enormous molecule. As the work of purifying the catalyst responsible for DNA synthesis proceeded, the scale of operations increased to the point where almost every week or so we must work up a hundred pounds of these bacteria to get enough of this trace catalyst to study in greater detail.

Now the sum total of these studies can be stated as follows--that given these four building blocks (A, G, T and C) and this enzyme, this catalyst extracted from bacteria or another source if you will, and DNA as a model or a template of a sort, DNA synthesis proceeds at a rapid rate and to the point of exhaustion of the building blocks that you supply in the test tube. The characteristics of the product that are synthesized in the test tube are indistinguishable from those that we recognise for DNA isolated from nature, and the way in which this happens can best be shown in the last slide.

We deduce from these studies that synthesis proceeds in this manner. Imagine that you have a small segment of DNA. These are four units out of the many thousands in a chain of DNA and for convenience we'll say that one strand has an A, C, T and G and of necessity the sister strand has T for A, G for C, A for T, C for G. When synthesis starts (preceding mitosis in the cell) the DNA must be reproduced. The strand on the left comes apart from the strand on the right and having separated them, let's picture the strand on the left--it's the same one--G, T, C and A; and coming out of the surrounding solution are the building blocks and for the G we demand a C and that fits; and the T calls for an A, and then once it's there, it hooks up with a C, in this cursive writing so to speak, of DNA. The C calls for a G, the A for a T. Now while this is going on, back at the ranch, down on the right hand side, units are being assembled to match the right hand strand; the C demands a G, T fits with A, C with G, A with T, and if you step back and look at this, you can see that each of

the daughter molecules is identical to the other. They have the same sequence and they represent a perfect replica of the parent In this device for replicating DNA which in fact was first suggested in Watson & Crick's model when they came upon it, is in fact the way in which DNA is assembled in the test tube.

As we have learned some of the details of DNA synthesis carried out under these controlled conditions, we have become painfully aware of all the things we do not know about it. We are puzzled by the conditions that are necessary to take these strands apart in the cell because as you know under normal conditions in an adult cell the DNA is dormant. Only under conditions of abnormal growth, such as cancer or some special simulus, does the DNA suddenly start to replicate in adult life. And we're concerned about the fact that in the test tube this synthesis goes on to the point of exhaustion of the building blocks whereas in the cell, it exactly doubles and stops. These points of control which still escape us represent important areas of future research.

We are also interested in whether by more careful control of laboratory synthesis of DNA, we can actually make the unit that is so perfect that it does represent a gene and can be inserted into a cell and transmit genetic information When this is accomplished, and I think it is within sight, if you have long range vision, then you can imagine inserting fraudulent letters and observing subsequent alterations in the message that is carried to the cell in question.

I might summarize what I have been telling you in a few general statements. I think it is clear, or I have tried to make it clear to you, that from every point of view, biology and medicine are coming closer to chemical interpretations of the phenomena that we study in these areas. In studying DNA we are studying the master code itself. Condensed in its atomic language, we have the directions for protein synthesis and the proteins are the molecules that govern all the cellular processes. And we recognize that there must be a mechanism for replicating DNA, perhaps like the one I have described to you, which is responsible for the preservation of the

race and if a mistake is made in the DNA or imposed on it, by radiation or chemicals that are mutagenic, these changes are of course copied and result in proteins which may be defective. This may be sufficiently great an alteration to produce death, and then of course that's the end of that. But if such a change produces a defect that does not cause death, we then have a mutant, we have an altered species which, if survival permits, keeps on going. Very rarely such an alteration may result in a favorable change under the particular conditions. So we can understand the very slow evolutionary change in biology and this kind of origin of species.

As we study nucleic acids and proteins in more detail, we come to a more aesthetic appreciation of biologic phenomena and we have the optimism and confidence that as we know this better, practical rewards in an understanding of disease inevitably will follow.

QUESTIONS

There is a bit of time... Dr Kornberg would be most happy to answer questions before he comes and asks me the questions I know he will--this is no idle threat about coming to a budget session bearing in hand what I said at the beginning.

I think those of you in the audience this evening, who are students, fully appreciate that there is lots of room at the bottom in Biology as there is in Electronics and the frontiers on which Dr. Kornberg and his associates are working are by no means conquered as yet. There's lots of room for all of you to work as well. Questions?

A gene until recently has been an abstract definition. Now it is embarrassing that we know what the genetic substance is and we've got to define it in more precise chemical terms and we are at a loss to do this precisely. We believe that a gene may be several thousand nucleotides--a stretch of them--and they are translated into a specific protein or enzyme and that is the gene. We do not know the punctuation in a DNA chain. We do not know, as I said a moment ago, the sequential arrangement of nucleotides, so that beyond recognising that a chromosome is probably an extremely long stretch of these individual nucleotides, punctuated at various points, as each gene follows another, we really can't give you any more structural detail--going from the molecular level up to the chromosome itself. I should say that in viruses which are so much simpler and don't have an obvious chromosome, just have DNA, the newest information indicates that where you have perhaps 200,000 nucleotides in this virus (and it was formerly thought that you might have ten separate DNA molecules) that with sufficient care in opening these viruses, it is actually a single molecule with all 200,000 nucleotides linked in a row. And we imagine that perhaps chromosomes are in fact just single chains of DNA, although this is not established